

transistors, a common mode voltage is applied to both pairs of transistors and a minimum input signal needed to control the output via the pair of small transistors is measured.

A measurement was made to characterize the percentage of times the output was correctly controlled by the input over 1000 cycles using a "001100. . ." input pattern. With a maximum common mode voltage (V_{common}) fixed to 0.9 V (100–200 mV above V_t) the statistical output as a function of delta input voltage (V_{input}) between the two "small" transistors is measured. FIG. 10 shows a loss of precision starting at $V_{input}=120$ mV. The correspondence between this measurement and precision depends critically on the precise threshold voltage of the devices; for a process with V_t varying between 0.7 V and 0.8 V which in the worst case means 8 bits and in the best case 9 bits of precision. This analysis is based on drawn transistor dimensions making exact precision characterization following fabrication difficulty.

Another test structure is realized to estimate the computing precision of the neuron comparator in FIG. 9. The test structure contains 1 k flash devices ($0.81 \mu\text{m}/2 \mu\text{m}$) for common mode input whose inputs are controlled by a shift register and 32 flash devices for differential input whose inputs are controlled by a latch. The test structure circuit also contains programming and erasing drivers for the flash devices. The entire 1 k flash devices represents the common mode signal for the positive and the negative synapses; 16 compensation flash devices are able to compensate the programming error and the remaining 16 input flash devices are used to apply the input signal to measure the overall precision. After all the flash devices have been programmed to $V_t=2.5$ V, the shift register is filled with "1" s. A gate voltage is fixed for all the flash devices to $V_t + V_{LSB}$. Applying a $V_t + V_{LSB}$ voltage to the flash gate devices means fixing a common mode signal for the positive and the negative synapses (as has been done by the big transistors of the previous test structure). After compensation, a sequence of "001 10011. . ." is applied to 1 to 16 of the inputs testing for the correct output over 1 million cycles.

FIG. 11 shows the percentage of correct output as a function of the number of inputs applied (1 to 16). It is clear that as the number of inputs increases the percentage of correct outputs increases, approaching 100% for 6 inputs with a $V_{LSB}=32$ mV. This corresponds to a precision of 7+ bits. FIG. 13 shows the precision as a function of the LSB voltage. As can be seen, increasing the LSB voltage to 128 mV, 1 bit of precision is lost. The effective precision of the circuit depends on the common mode signal. FIG. 12 shows the same test using only 0.5 k flash devices (reduced common mode) In this case, the differential input needed for 100% correct output is 2 synapses 15 or 17 being on and the effective precision is close to 8 bits.

FIG. 14 shows a measurement of the neuron functionality clocked at 10 MHz. FIG. 15 shows power consumption of the neuron over one computation. The consumption is equal to 166 pJ at 5 V of power supply at a frequency of 1.7 MHz. This corresponds to 166 pJ per input multiply-accumulate operation (1 k inputs).

Clearly, changes may be made to the neural network as described and illustrated herein without, however, departing from the scope of the present invention. A number of alternative embodiments have been described herein for implementing a conductance synapse in a neuron. The invention is to be interpreted broadly, covering all embodiments and modifications thereof which fall within the scope of the appended claims.

What is claimed is:

1. A neural network comprising:

a plurality of synaptic weighting elements organized in a first set and a second set, each having a respective programmable conductance, each of the synaptic weighting elements having a respective synaptic input connection for receiving an input signal and an output connection;

a neuron stage coupled to the synaptic weighting elements;

a conductance comparing circuit within the neuron stage, the conductance comparing circuit comparing the conductance of the first set of synaptic weighting elements to the conductance of the second set of synaptic weighting elements and outputting an indication of which set has the higher conductance without determining the conductance values; and

a latch stage coupled to the neuron stage to digitize the outputted indication.

2. The network according to claim 1 wherein the synaptic weighting elements include a nonvolatile memory cell having a floating gate and a control gate in a programmable conductance.

3. The network according to claim 1, further including:

an exciting conductive input terminal coupled to the conductance comparing circuit;

an inhibiting conductance input terminal coupled to the conductance comparing circuit;

a connection between the first set of the synaptic elements and the exciting conductive input terminal; and

a connection between the second set of synaptic elements and the inhibiting conductive input terminal, the conductance comparing circuit outputting an indication of the results of the comparison between the exciting synaptic elements and the inhibiting synaptic elements.

4. A neural network comprising:

a plurality of exciting synaptic elements having a programmable conductance;

a plurality of inhibiting synaptic elements having a programmable conductance;

a neuron stage having a first and second conductance input connected to the exciting and inhibiting synaptic elements, respectively, the neuron stage including a comparing circuit for providing a comparison of the conductance of the inhibiting synaptic elements and the exciting synaptic elements and generating an output signal indicative of which conductance is greater without determining the value of the conductance of the inhibiting synaptic elements and the exciting synaptic elements; and

a latch stage coupled to the neuron stage to digitize the output signal.

5. A neural network comprising:

a plurality of exciting and inhibiting synaptic weighting elements having a respective programmable conductance;

a neuron stage coupled to the plurality of exciting and inhibiting synaptic weighting elements, each of the synaptic weighting elements having an associated exciting or inhibiting conductance and an input connection terminal for receiving a respective input signal and an output coupled to the neuron stage, the neuron stage further including a conductance sensing circuit that generates an output signal indicative of which of the total exciting and inhibiting conductances of the plurality of synaptic weighting elements is greater

13

without determining the actual value, collectively and individually, of the exciting and inhibiting conductances; and

a latch stage coupled to the neuron stage to digitize the output signal.

6. A neural network comprising:

a plurality of exciting synaptic weighting elements, each having a respective programmable conductance, each of the exciting synaptic weighting elements having output terminals which are connected in common to each other to provide an output which is the sum of the conductances of all the exciting synaptic weighting elements;

a plurality of inhibiting synaptic weighting elements, each having a respective programmable conductance, each of the inhibiting synaptic weighting elements having output terminals which are connected in common to each other to provide an output terminal having an

14

output which is the sum of the conductances of all the inhibiting synaptic weighting elements;

a neuron stage having an exciting input connected to the output of the plurality of exciting synaptic weighting elements and having an inhibiting input connected to the output of the plurality of the inhibiting synaptic weighting elements; and

a conductance comparing stage for comparing the conductance between the sum of the exciting synaptic weighting elements and the inhibiting synaptic weighting elements without determining the value of the conductances and outputting a signal indicative of which of the sums of the respective pluralities of synaptic weighting elements has the greatest conductance.

* * * * *